

WHAT IS CLAIMED IS:

1.    A method of creating a searchable archive accessible by a  
5    data processing system, comprising:

         generating a domain structure and tokenized data  
         from an archive data set, the domain structure including  
         tokens corresponding to unique values in the archive data  
         set and the tokenized data including token columns  
10    corresponding to value columns in the archive data set;

         determining archive metadata from the domain  
         structure and the tokenized data;

         dividing the tokenized data into one or more token  
         column segments;

15    determining token column segment metadata from the  
         one or more token column segments;

         creating one or more compressed token column  
         segments from the token column segments;

         creating one or more compacted files from the one or  
20    more compressed token column segments and the token  
         column segment metadata; and

         storing the one or more compacted files in a file  
         system coupled to the data processing system.

25    2.    The method of claim 1, wherein determining metadata  
         further comprises determining a maximum value and a minimum  
         value for each of the token columns.

         3.    The method of claim 1, wherein determining metadata  
30    further comprises determining a maximum tupleid and a minimum  
         tupled for each of the one or more token column segments.

         4.    The method of claim 1, further comprising:  
                dividing the domain structure into one or more  
35    domain structure segments;

1       **51681/FLC/S673**

          determining metadata from the domain structure  
segments;

5       compressing the one or more domain structure  
segments; and

          creating one or more compacted files further  
includes storing the compressed domain structure segments  
in the compacted file.

10

5.   A method of retrieving a datum from a searchable archive  
by a data processing system, the searchable archive comprising  
a metadata file and one or more compacted files, comprising:

15       selecting a selected compacted file from the one or  
more compacted files that may include the datum using the  
metadata file;

          accessing the selected compacted file;

20       selecting a selected compressed segment from one or  
more compressed segments in the selected compacted file  
using metadata stored in the compacted file;

          generating a decompressed segment from the selected  
compressed segment; and

25       searching the decompressed segment to determine if  
the decompressed segment includes the datum.

6. The method of claim 5 wherein:

          selecting a selected compacted file is performed by  
a search process; and

30       accessing the selected compacted file, selecting a  
selected compressed segment, generating a decompressed  
segment, and searching the decompressed segment are  
performed by one or more search agents invoked by the  
search process.

35

5       7.    A method of creating a searchable archive accessible by a  
data processing system, comprising:

          generating a domain structure and tokenized data  
          from archive data;

          determining metadata from the tokenized data;

10       generating a set of bit vectors from the tokenized  
data;

          creating one or more compacted files from the set of  
bit vectors; and

          storing the one or more compacted files in a file  
system coupled to the data processing system.

15

8.    The method of claim 7, wherein the tokenized data set  
includes one or more columns of tokens and extracting archive  
metadata further comprises determining a maximum token value  
and a minimum token value for each of the one or more columns  
20    of tokens.

9.    A method of retrieving a datum from a searchable archive  
by a data processing system, the searchable archive comprising  
a metadata file and one or more compacted files, comprising:

25       selecting a selected compacted file from the one or  
more compacted files that may include the datum using the  
metadata;

          accessing the selected compacted file;

30       selecting one or more bit vectors from the selected  
compacted file; and

          performing a Boolean operation on the bit vectors  
included in the selected compacted file to determine if  
the datum is stored in the selected compacted file.

35    10. The method of claim 9, wherein:

selecting a selected compacted file is performed by  
a search process; and

5           accessing the selected compacted file and performing  
a Boolean operation is performed by one or more search  
agents invoked by the search process.

11. A data processing system for creating a searchable  
10 archive, comprising:

          a processor; and

          a memory coupled to the processor, the memory having  
program instructions executable by the processor stored  
therein, the program instructions including:

15           generating a domain structure and tokenized  
data from an archive data set, the domain structure  
including tokens corresponding to unique values in  
the archive data set and the tokenized data  
including token columns corresponding to value  
20 columns in the archive data set;

          determining archive metadata from the domain  
structure and the tokenized data;

          dividing the tokenized data into one or more  
token column segments;

25           determining token column segment metadata from  
the one or more token column segments;

          creating one or more compressed token column  
segments from the token column segments;

          creating one or more compacted files from the  
one or more compressed token column segments and the  
token column segment metadata; and  
30

          storing the one or more compacted files in a  
file system coupled to the data processing system.

35   12. The data processing system of claim 11, the program

1       **51681/FLC/S673**

instructions for determining metadata further including  
determining a maximum value and a minimum value for each of  
5       the token columns.

13. The data processing system of claim 11, the program  
instructions for determining metadata further including  
determining a maximum tupleid and a minimum tupled for each of  
10       the one or more token column segments.

14. The data processing system of claim 11, the program  
instructions further including:

15       dividing the domain structure into one or more  
domain structure segments;

          determining metadata from the domain structure  
segments;

          compressing the one or more domain structure  
segments; and

20       creating one or more compacted files further  
includes storing the compressed domain structure segments  
in the compacted file.

15. A data processing system for retrieving a datum from a  
25       searchable archive, the searchable archive comprising a  
metadata file and one or more compacted files, comprising:

          a processor; and

          a memory coupled to the processor, the memory having  
program instructions executable by the processor stored  
30       therein, the program instructions including:

          selecting a selected compacted file from the  
one or more compacted files that may include the  
datum using the metadata file;

          accessing the selected compacted file;

35       selecting a selected compressed segment from

1       **51681/FLC/S673**

          one or more compressed segments in the selected  
          compacted file using metadata stored in the  
5       compacted file;  
          generating a decompressed segment from the  
          selected compressed segment; and  
          searching the decompressed segment to determine  
          if the decompressed segment includes the datum.

10

16. The data processing system of claim 15, the program  
instructions further including:

          selecting a selected compacted file is performed by  
          a search process; and  
15       accessing the selected compacted file, selecting a  
          selected compressed segment, generating a decompressed  
          segment, and searching the decompressed segment are  
          performed by one or more search agents invoked by the  
          search process.

20

17. A data processing system for creating a searchable  
archive, comprising:

          a processor; and  
          a memory coupled to the processor, the memory having  
25       program instructions executable by the processor stored  
          therein, the program instructions including:  
          generating a domain structure and tokenized  
          data from archive data;  
          determining metadata from the tokenized data;  
30       generating a set of bit vectors from the  
          tokenized data;  
          creating one or more compacted files from the  
          set of bit vectors; and  
          storing the one or more compacted files in a  
35       file system coupled to the data processing system.

5       18. The data processing system of claim 17, wherein the  
tokenized data set includes one or more columns of tokens, the  
program instructions for extracting archive metadata further  
including determining a maximum token value and a minimum  
token value for each of the one or more columns of tokens.

10      19. A data processing system for retrieving a datum from a  
searchable archive, the searchable archive comprising a  
metadata file and one or more compacted files, comprising:

          a processor; and  
          a memory coupled to the processor, the memory having  
15      program instructions executable by the processor stored  
therein, the program instructions including:

          selecting a selected compacted file from the  
          one or more compacted files that may include the  
          datum using the metadata;

20      accessing the selected compacted file;

          selecting one or more bit vectors from the  
          selected compacted file; and

          performing a Boolean operation on the bit  
          vectors included in the selected compacted file to  
25      determine if the datum is stored in the selected  
          compacted file.

20. The data processing system of claim 19, wherein:

30      selecting a selected compacted file is performed by  
a search process; and

          accessing the selected compacted file and performing  
a Boolean operation is performed by one or more search  
agents invoked by the search process.

35

21. A method of utilizing a searchable archive by a data  
5 processing system, comprising:

      generating a domain structure and tokenized data  
      from archive data;

      determining archive metadata from the tokenized  
      data;

10       dividing the tokenized data into one or more  
      segments;

      determining segment metadata from the one or more  
      segments;

      creating one or more compressed segments from the  
15 segments;

      creating one or more compacted files from the one or  
      more compressed segments and the segment metadata; and

      storing the one or more compacted files in a file  
      system coupled to the data processing system.

20       22. The method of claim 21, further comprising:

      selecting a selected compacted file from the one or  
      more compacted files that may include a datum using the  
      archive metadata;

25       accessing the selected compacted file;

      selecting a selected compressed segment from the one  
      or more compressed segments in the selected compacted  
      file using the segment metadata;

      generating a decompressed segment from the selected  
30 compressed segment; and

      searching the decompressed segment to determine if  
      the decompressed segment includes the datum.

23. The method of claim 22 wherein:

35       selecting a selected compacted file is performed by



a search process; and

5        accessing the selected compacted file, selecting a  
selected compressed segment, generating a decompressed  
segment, and searching the decompressed segment are  
performed by one or more search agents invoked by the  
search process.

10      24. A method of utilizing a searchable archive by a data  
processing system, comprising:

generating a domain structure and tokenized data  
from archive data;

15      determining archive metadata from the tokenized  
data;

generating a set of bit vectors from the tokenized  
data;

creating one or more compacted files from the set of  
bit vectors; and

20      storing the one or more compacted files in a file  
system coupled to the data processing system.

25. The method of claim 24, further comprising:

25      selecting a selected compacted file from the one or  
more compacted files that may include a datum using the  
archive metadata;

accessing the selected compacted file;

selecting one or more bit vectors from the selected  
compacted file; and

30      performing a Boolean operation on the bit vectors  
included in the to determine if the datum is stored in  
the compacted file.

26. The method of claim 25, wherein:

35      selecting a selected compacted file is performed by

a search process; and  
          accessing the selected compacted file and  
5        performing a Boolean operation is performed by one  
          or more search agents invoked by the search process.

27. A data processing system for utilizing a searchable  
archive, comprising:

10        a processor; and  
          a memory coupled to the processor, the memory having  
          program instructions executable by the processor stored  
          therein, the program instructions including:  
          generating a domain structure and tokenized  
15        data from archive data;  
          determining archive metadata from the tokenized  
          data;  
          dividing the tokenized data into one or more  
          segments;  
20        determining segment metadata from the one or  
          more segments;  
          creating one or more compressed segments from  
          the segments;  
          creating one or more compacted files from the  
25        one or more compressed segments and the segment  
          metadata; and  
          storing the one or more compacted files in a  
          file system coupled to the data processing system.

30        28. The data processing system of claim 27, the program  
          instructions further including:

          selecting a selected compacted file from the one or  
          more compacted files that may include a datum using the  
          archive metadata;  
35        accessing the selected compacted file;

1       **51681/FLC/S673**

          selecting a selected compressed segment from the one  
or more compressed segments in the selected compacted  
5       file using the segment metadata;

          generating a decompressed segment from the selected  
compressed segment; and

          searching the decompressed segment to determine if  
the decompressed segment includes the datum.

10

29. The data processing system of claim 28, wherein

          selecting a selected compacted file is performed by  
a search process; and

15

          accessing the selected compacted file, selecting a  
selected compressed segment, generating a decompressed  
segment, and searching the decompressed segment are  
performed by one or more search agents invoked by the  
search process.

20

30. A data processing system for utilizing a searchable  
archive, comprising:

          a processor; and

          a memory coupled to the processor, the memory having  
program instructions executable by the processor stored  
25       therein, the program instructions including:

25

          generating a domain structure and tokenized  
data from archive data;

          determining archive metadata from the tokenized  
data;

30

          generating a set of bit vectors from the  
tokenized data;

          creating one or more compacted files from the  
set of bit vectors; and

35

          storing the one or more compacted files in a  
file system coupled to the data processing system.

1

**51681/FLC/S673**

5

31. The data processing system of claim 30, the program instructions further including:

selecting a selected compacted file from the one or more compacted files that may include a datum using the archive metadata;

accessing the selected compacted file;

10

selecting one or more bit vectors from the selected compacted file; and

performing a Boolean operation on the bit vectors included in the to determine if the datum is stored in the compacted file.

15

32. The data processing system of claim 31, wherein:

selecting a selected compacted file is performed by a search process; and

accessing the selected compacted file and

20

performing a Boolean operation is performed by one or more search agents invoked by the search process.

25

30

35